

REL Southwest Ask A REL Response

January 2020

Question:

Is there a body of research that validates the practice of using interim assessments as a way to improve student outcomes on end-of-year summative assessments?

Response:

Thank you for the question you submitted to our REL Reference Desk. We have prepared the following memo with research references to help answer your question. For each reference, we provide an abstract, excerpt, or summary written by the study's author or publisher. Following an established Regional Educational Laboratory (REL) Southwest research protocol, we conducted a search for research on the relationship between interim assessments and improving student outcomes on summative assessments.

We have not evaluated the quality of references and the resources provided in this response. We offer them only for your reference. Also, we searched the references in the response from the most commonly used resources of research, but they are not comprehensive, and other relevant references and resources may exist. References provided are listed in alphabetical order, not necessarily in order of relevance. We do not include sources that are not freely available to the requestor.

Research References

Babo, G., Tienken, C. H., & Gencarelli, M. A. (2014). Interim testing, socio-economic status, and the odds of passing grade 8 state tests in New Jersey. *RMLE Online: Research in Middle Level Education*, 38(3), 1–9. <https://eric.ed.gov/?id=EJ1047113>

From the ERIC abstract: “A review of the literature pertaining to the effect and influence of commercially-prepared interim assessments in mathematics and language arts literacy reveals a lack of quantitative data to determine the value of these products for school reform. This study examined the ability of commercially-prepared interim pretest and posttest assessments in language arts literacy (LAL) and math to predict student achievement on the state-mandated summative assessment in those subjects. Analyses were conducted using binary logistic regression models. Data for this study included results from the state-mandated grade 8 assessments (NJ ASK 8) for 291 eighth grade students enrolled in two middle level schools located in a suburban/urban central New

Jersey community during the 2009-2010 academic school year. The findings suggest that the predictive value of the students' pretest results is very similar to that of the posttest results and call into question the efficacy of implementing both interim pretests and posttests."

Blanc, S., Christman, J. B., Liu, R., Mitchell, C., Travers, E., & Bulkley, K. E. (2010). Learning to learn from data: Benchmarks and instructional communities. *Peabody Journal of Education*, 85(2), 205–225. <https://eric.ed.gov/?id=EJ883193>. Retrieved from <https://www.researchgate.net/publication/248942488>

From the ERIC abstract: "This article examines the use of interim assessments in elementary schools in the School District of Philadelphia. The article reports on the qualitative component of a multimethod study about the use of interim assessments in Philadelphia. The study used an organizational learning framework to explore how schools can best develop the capacity to utilize the potential benefits of interim assessments. The qualitative analysis draws on data from intensive fieldwork in 10 elementary schools and interviews with district staff and others who worked with the schools, as well as further in-depth case study analysis of 5 schools. This article examines how school leaders and grade groups made sense of data provided through interim assessments and how they were able to use these data to rethink instructional practice. We found substantial evidence that interim assessments have the potential to contribute to instructional coherence and instructional improvement if they are embedded in a robust feedback system. Such feedback systems were not the norm in the schools in our study, and their development requires skill, knowledge, and concerted attention on the part of school leaders."

Bulkley, K. E., Christman, J. B., Goertz, M. E., & Lawrence, N. R. (2010). Building with benchmarks: The role of the district in Philadelphia's benchmark assessment system. *Peabody Journal of Education*, 85(2), 186–204. <https://eric.ed.gov/?id=EJ883188>. Retrieved from <https://www.researchgate.net/publication/233120636>

From the ERIC abstract: "In recent years, interim assessments have become an increasingly popular tool in districts seeking to improve student learning and achievement. Philadelphia has been at the forefront of this change, implementing a set of Benchmark assessments aligned with its Core Curriculum district-wide in 2004. In this article, we examine the overall context for Benchmarks in Philadelphia, the expectations district leaders had for the use of those Benchmarks, the supports put in place to assist those in schools in meeting those expectations, and the challenges encountered in that implementation."

Chojnacki, G., Eno, J., Liu, F., Meyers, C., Konstantopoulos, S., Miller, S., & van der Ploeg, A. (2013, September). *Do interim assessments influence instructional practice in year one? Evidence from Indiana elementary school teachers*. Paper presented at the Society for Research on Educational Effectiveness Fall 2013 Conference, Washington, District of Columbia. <https://eric.ed.gov/?id=ED563059>

From the ERIC abstract: “Recent work that has examined the impact of what are variously called periodic, interim, benchmark, or diagnostic assessments, typically administered three or four times during a school year, has produced mixed findings. For instance, one study reported small significant effects in mathematics in grades 3-8, but not in reading (Carlson et al., 2011). Other research however, has reported significant effects on both mathematics and reading (Slavin et al., 2011). Finally, a very recent study found no effects on reading achievement in grades 4-5 (Cordray et al., 2012). This study compares instructional practices of teachers in schools that were randomly assigned to receive an interim assessment tool with those of teachers in schools that did not receive the tool. Using rich data collected at 16 time points during the school year, the authors study teachers’ self-reported instructional practices to determine whether teachers with access to an interim assessment tool alter each of three facets of instructional practice—scope and sequence of content coverage, instructional level, and instructional grouping—more than those without the tool. The research questions are: (1) Do teachers with access to the interim assessment change the scope and sequence of content, and/or vary instructional difficulty level and grouping methods more than those without? (2) Do variations in these teacher practices respond to variations in student Acuity performance? Researchers employ treatment vs. control comparisons to explore whether teachers with the interim assessment intervention engage in expected instructional practices more than those without it. Results are reported from rich data on teacher instructional practices generated at sixteen intervals by teachers with and without access to a specific interim assessment tool. Estimates provide no strong evidence that teachers change the instructional practices measured here in response to Acuity performance data. One possible reason for these findings is that Acuity is not a unique intervention, and a significant number of control teachers reported using other interim assessment tools. Another possible explanation for these results is that the relatively small sample of teachers completing checklists harms the study’s power. Finally, these results pertain to the first year of the intervention, when teachers are likely still learning how to use the assessment tool and integrate it into their instructional practice. Future research should explore the hypothesis that impacts on teacher practice grow over time as teachers learn to use the assessment tool.”

Cordray, D., Pion, G., Brandt, C., Molefe, A., & Toby, M. (2013). *The impact of the Measures of Academic Progress (MAP) program on student reading achievement* (Final Report, NCEE 2013-4000). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://eric.ed.gov/?id=ED537982>

From the ERIC abstract: “During the past decade, the use of standardized benchmark measures to differentiate and individualize instruction for students received renewed attention from educators. Although teachers may use their own assessments (tests, quizzes, homework, problem sets) for monitoring learning, it is challenging for them to equate performance on classroom measures with likely performance on external measures, such as statewide tests or nationally normed standardized tests. One of the most widely used commercially available systems incorporating benchmark assessment and training in differentiated instruction is the Northwest Evaluation Association’s (NWEA) Measures of Academic Progress (MAP) program. The MAP program includes:

(1) computer-adaptive assessments administered to students three or four times a year; and (2) teacher training and access to MAP resources on how to use data from these assessments to differentiate instruction. MAP tests and training are currently in use in nearly 20 percent of K-12 school districts nationwide and more than a third of districts in the Midwest. Although the technical merits and popularity of MAP assessments have been widely referenced in practitioner-oriented journals and teacher magazines, few studies have investigated the effects of MAP or other benchmark assessment programs on student outcomes. This study was designed to address questions from Midwestern states and districts about the extent to which benchmark assessment may affect teachers' differentiated instructional practices and student achievement. Thirty-two elementary schools in five districts in Illinois participated in a two-year randomized controlled trial to assess the effectiveness of the MAP program. Half the schools were randomly assigned to implement the MAP program in grade 4, and the other half were randomly assigned to implement MAP in grade 5. Schools assigned to grade 4 treatment served as the grade 5 control condition, and schools assigned to grade 5 treatment served as the grade 4 control. The results of the study indicate that the MAP program was implemented with moderate fidelity but that MAP teachers were not more likely than control group teachers to have applied differentiated instructional practices in their classes. Overall, the MAP program did not have a statistically significant impact on students' reading achievement in either grade 4 or grade 5."

Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E., Supovitz, J. A., & Wayman, J. C. (2009). *Using student achievement data to support instructional decision making* (IES Practice Guide, NCEE 2009-4067). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://eric.ed.gov/?id=ED506645>

From the ERIC abstract: "The purpose of this practice guide is to help K-12 teachers and administrators use student achievement data to make instructional decisions intended to raise student achievement. The panel believes that the responsibility for effective data use lies with district leaders, school administrators, and classroom teachers and has crafted the recommendations accordingly. This guide focuses on how schools can make use of common assessment data to improve teaching and learning. For the purpose of this guide, the panel defined common assessments as those that are administered in a routine, consistent manner by a state, district, or school to measure students' academic achievement. These include: (1) annual statewide accountability tests such as those required by No Child Left Behind; (2) commercially produced tests—including interim assessments, benchmark assessments, or early-grade reading assessments—administered at multiple points throughout the school year to provide feedback on student learning; (3) end-of-course tests administered across schools or districts; and (4) interim tests developed by districts or schools, such as quarterly writing or mathematics prompts, as long as these are administered consistently and routinely to provide information that can be compared across classrooms or schools. This guide includes five recommendations that the panel believes are a priority to implement: (1) Make data part of an ongoing cycle of instructional improvement; (2) Teach students to examine their own data and set learning goals; (3) Establish a clear vision for schoolwide data use; (4) Provide supports

that foster a data-driven culture within the school; and (5) Develop and maintain a districtwide data system.”

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2007). *Measuring how benchmark assessments affect student achievement* (Issues & Answers Report, REL 2007-039). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. <https://eric.ed.gov/?id=ED499792>

From the ERIC abstract: “This report examines a Massachusetts pilot program for quarterly benchmark exams in middle-school mathematics, finding that program schools do not show greater gains in student achievement after a year. But that finding might reflect limited data rather than ineffective benchmark assessments. Benchmark assessments are used in many districts throughout the nation to raise student, school, and district achievement and to meet the requirements of the No Child Left Behind Act of 2001. This report details a study using a quasi-experimental design to examine whether schools using quarterly benchmark exams in middle-school mathematics under a Massachusetts pilot program show greater gains in student achievement than schools not in the program. The following are appended: (1) Methodology; (2) Construction of the Study Database; (3) Identification of Comparison Schools; (4) Interrupted Time Series Analysis; and (5) Massachusetts Curriculum Frameworks for Grade 8 Mathematics (May 2004).”

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2008). *A second follow-up year for “Measuring how benchmark assessments affect student achievement.” REL Technical Brief* (REL 2008-002). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. <https://eric.ed.gov/?id=ED501327>

From the ERIC abstract: “This technical brief examines whether, after two years of implementation, schools in Massachusetts using quarterly benchmark exams aligned with state standards in middle school mathematics showed greater gains in student achievement than those not doing so. A quasi-experimental design, using covariate matching and comparative interrupted time-series techniques, was used to assess school differences in changes in mathematics performance between program and comparison schools. Following up on an earlier report with just one year of post-implementation data, the study found no significant differences between schools using this practice and those not doing so after two years. The brief summarizes findings from a follow-up study to the Issues & Answers report, ‘Measuring How Benchmark Assessments Affect Student Achievement. REL 2007-No. 039’ [ED499792]. The follow-up study adds another year of post-implementation data to examine the impact of benchmark assessments on grade 8 mathematics achievement, using the same data sources, methods, and reporting as the original study. The study examines whether, after two years of implementation, schools in Massachusetts using quarterly benchmark exams aligned with state standards in middle school mathematics showed greater gains in student achievement than those not doing so. A quasi-experimental design, using covariate matching and comparative interrupted time-series techniques, was used to assess differences in changes in mathematics performance

between program and comparison schools. The follow-up study finds no significant differences between schools using this practice and those not doing so after two years. Limitations include the lack of data on what benchmark assessment practices comparison schools may be using, having only 22 treatment and 44 comparison schools, and having only two years of post-implementation data—perhaps still too few to observe an impact from the intervention.”

Konstantopoulos, S., Li, W., Miller, S., & van der Ploeg, A. (2015, March). *Effects of interim assessments on the achievement gap: Evidence from an experiment*. Paper presented at the Society for Research on Educational Effectiveness Spring 2015 Conference, Washington, District of Columbia. <https://eric.ed.gov/?id=ED562166>

From the ERIC abstract: “Motivated by the passage of the No Child Left Behind (NCLB) Act, all states operate accountability systems that measure and report school and student performance annually. The purpose of this study is to examine the effects of interim assessments on the achievement gap. The authors examine the impact of interim assessments throughout the distribution of student achievement with a focus on the lower tail of the achievement distribution. Specifically, they investigated the effects of two interim assessment programs (i.e., ‘mCLASS’ and ‘Acuity’) on mathematics and reading achievement for high- median- and low-achievers. They use data from a large-scale experiment conducted in the state of Indiana in the 2009-2010 school year. Quantile regression is used to analyze student data. The study was a large-scale experiment conducted in Indiana during the 2009-2010 academic year and included K-8 public schools that had volunteered to participate in the intervention in the spring of 2009. From a stratified (by school urbanicity) pool of 116 schools the authors randomly selected 70 schools. Ten of the 70 schools had used one or both assessment programs the prior year and were excluded from the pool. Two other schools closed and another school did not provide any student data. Thus, the final sample included 57 schools, 35 in treatment and 22 in control condition. Overall, nearly 20,000 students participated in the study during the 2009-2010 school year. The design was a two-level cluster randomized design. Students were nested within schools, and schools were nested within treatment and control conditions. Schools were randomly assigned to a treatment (interim assessment) or a control condition. The schools in the treatment condition received ‘mCLASS’ and ‘Acuity’, and the training associated with each program. The control schools operated under business-as-usual conditions. Overall, the findings suggest that the treatment effect was positive, but not consistently significant across all grades. Significant treatment estimates were observed in the grade 3-8 analysis in mathematics. The estimates were typically larger for low-achievers and in some cases significant. These results are consistent in terms of the sign of the effect (i.e., positive), but inconsistent in terms of statistical significance. The authors observed positive, statistically significant effects for grades 3-8 especially in mathematics. It seems that ‘Acuity’ affected mathematics and reading achievement positively and in some instances considerably in grades 3-6.”

Pereira, M., & Tienken, C. (2012). An evaluation of the influence of interim assessments on grade 8 student achievement in mathematics and language arts. *International Journal of Educational Leadership Preparation*, 7(3), 1–13. <https://eric.ed.gov/?id=EJ997471>

From the ERIC abstract: “A review of the literature pertaining to the effect and influence that interim assessments have on student achievement lacks quantitative data to determine the efficiency of their use in the classroom as a school reform tool. This study examined the strength and the direction of the relationships between interim pre and posttest assessments in language arts and mathematics in Grade 8 and student achievement on the New Jersey Grade 8 state standardized tests in those subjects. Analyses were conducted using simultaneous multiple regression models. All student data explored in this study pertained to 670 students in Grade 8 enrolled in four middle schools located in a suburban/urban central New Jersey community during the 2009-2010 academic school year. The results of the study revealed each school produced a combination of site specific results and the interim pretests accounted for the same or almost the same amount of variance in state test scores as the interim posttests.”

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5–13. <https://eric.ed.gov/?id=EJ853799>. Retrieved from <https://www.nciea.org/sites/default/files/pubs-tmp/Moving%20towards%20a%20comprehensive%20assessment%20system%20A%20framework%20for%20considering%20interim%20assessments.pdf>

From the ERIC abstract: “Local assessment systems are being marketed as formative, benchmark, predictive, and a host of other terms. Many so-called formative assessments are not at all similar to the types of assessments and strategies studied by Black and Wiliam (1998) but instead are interim assessments. In this article, we clarify the definition and uses of interim assessments and argue that they can be an important piece of a comprehensive assessment system that includes formative, interim, and summative assessments. Interim assessments are given on a larger scale than formative assessments, have less flexibility, and are aggregated to the school or district level to help inform policy. Interim assessments are driven by their purpose, which fall into the categories of instructional, evaluative, or predictive. Our intent is to provide a specific definition for these ‘interim assessments’ and to develop a framework that district and state leaders can use to evaluate these systems for purchase or development. The discussion lays out some concerns with the current state of these assessments as well as hopes for future directions and suggestions for further research.”

West, M. R., Morton, B. A., & Herlihy, C. M. (2016). *Achievement Network’s Investing in Innovation expansion: Impacts on educator practice and student achievement*. Cambridge, MA: Center for Education Policy Research, Harvard University. <https://eric.ed.gov/?id=ED565458>

From the ERIC abstract: “Data-based instructional programs have proliferated in American schools despite limited evidence of their effectiveness in improving educator practice and raising student achievement. We report results from a two-year school-randomized evaluation of the Achievement Network (ANet), a program providing schools with standards-aligned interim assessments and intensive supports for instructional data use. Survey data show that ANet increased teacher satisfaction with the timeliness and clarity of the data they receive and available supports for instructional

data-use and caused them to review and use interim assessment data more often. ANet did not, however, affect their confidence in data use or how frequently they differentiated instruction. Student impact estimates show no overall effect on student achievement in English language arts or mathematics. Despite the lack program effects on student achievement, we find that achievement is positively correlated with our survey-based measures of teacher perceptions and practices around instructional data use. Exploratory analyses suggest that the success of ANet in improving teacher practice and student achievement varies with the pre-existing capacity of schools to engage in data-based instruction. Schools rated by program staff as having a high level of readiness to implement the intervention prior to random assignment experienced positive impacts on student achievement, while those rated as having a low level of readiness experienced negative impacts.”

REL Southwest note: What Works Clearinghouse (WWC) rating: Meets Evidence Standards without Reservations. <https://ies.ed.gov/ncee/wwc/Study/84085>

What Works Clearinghouse. (2015). *WWC review of the report “The impact of Indiana’s system of interim assessments on mathematics and reading”* (What Works Clearinghouse Single Study Review). Washington, DC: U.S. Department of Education, Institute of Education Sciences. <https://eric.ed.gov/?id=ED553423>

From the ERIC abstract: “The study, ‘The Impact of Indiana’s System of Interim Assessments on Mathematics and Reading,’ examined the effects of using Diagnostic Assessment Tools (DAT) on mathematics and reading outcomes for students in 59 Indiana schools during the 2009-10 academic year. DAT consists of interim assessment tools—Wireless Generation’s mCLASS for students in grades K-2 and CTB/McGraw-Hill’s Acuity for students in grades 3-8-modified to align with Indiana’s state assessments. The goal is for teachers to use the assessment results to tailor instruction to students’ needs. After random assignment, schools in the intervention group received DAT, and schools in the comparison group did not receive the assessment tools or associated training. The study is a well-executed randomized controlled trial with low sample attrition. A subset of the analyses described in the study meets WWC group design standards without reservations. The study authors found, and the WWC confirmed, that the use of DAT did not have a statistically significant impact on general mathematics achievement or reading achievement for the full sample of students in grades K-8, but that the use of DAT did have statistically significant positive effects for grades 5 and 6 in mathematics achievement and grades 3-5 in reading achievement.”

REL Southwest note: WWC rating of the study reviewed: Meets Evidence Standards without Reservations.

Wilcox, K. C., Gregory, K., & Yu, L. (2017). Connecting the dots for English language learners: How odds-beating elementary school educators monitor and use student performance data. *Journal for Leadership and Instruction*, 16(1), 37–43. <https://eric.ed.gov/?id=EJ1159864>

From the ERIC abstract: “This article reports on findings from a multiple case study investigating the nature of educators’ approaches toward monitoring English language learners’ (ELLs) performance and using data to improve instruction and apply appropriate interventions. Six New York elementary schools where ELLs’ performance was better than predicted (i.e. odds-beating) based on student assessment data were studied. The analysis revealed that several strategies were common among the schools studied and were associated with the schools’ better ELL performance outcomes. These include: 1) connecting instruction and interventions to ‘real time’ data based on multiple measures of student performance including benchmark and formative assessments; 2) communicating performance via technology among teachers and with family members and legal guardians; 3) collaborating through routines among teaching and support staff as well as school and district leaders. Implications for district and school leaders and teachers are discussed. Implications for district and school leaders as well as teachers and other instructional specialists are offered.”

Additional Organizations to Consult

Center for Assessment – <https://www.nciea.org/>

From the website: “Comprehensive and balanced assessment systems are the subject of current technical and policy conversations, but designing effective and efficient systems can be fraught with major obstacles. Center professionals work with states and districts to first help identify highest priority uses and outline a Theory of Action. They then design and implement an assessment solution that may include formative, interim, and/or large-scale summative assessments, to meet the identified needs. In addition to their assessment expertise in general, Center professionals are recognized as national leaders in assessment design for students with significant cognitive disabilities and English language learners. The Center is also a leader in designing innovative assessment system to support educational reforms.”

REL Southwest note: The Center for Assessment provides relevant resources such as “The Marginal Utility of Interim Assessments for Prediction,” available at <https://www.nciea.org/library/marginal-utility-interim-assessments-prediction>, and “Using Interim Assessments in Place of Summative Assessments? Consideration of an ESSA Option,” available at <https://www.nciea.org/library/using-interim-assessments-place-summative-assessments-consideration-essa-option>.

Smarter Balanced Assessment Consortium – <http://smarterbalanced.org>

From the website: “Smarter Balanced is a public agency currently supported by its members. Through the work of thousands of educators, we created an online assessment system aligned to the Common Core State Standards (CCSS), as well as tools for educators to improve teaching and learning. Smarter Balanced is housed at the University of California Santa Cruz Silicon Valley Extension.

Our work is guided by the belief that a high-quality assessment system can provide information and tools for teachers and schools to improve instruction and help students

succeed—regardless of disability, language, or subgroup. We involve experienced educators, researchers, state and local policymakers, and community groups working together in a transparent and consensus-driven process.”

REL Southwest note: Smarter Balanced provides an overview of interim assessments. available at <https://portal.smarterbalanced.org/library/en/interim-assessments-overview.pdf>.

The Center on Standards and Assessment Implementation (CSAI) – <https://www.csai-online.org>

From the website: “The nation faces an unprecedented education challenge as nearly all of our states work to implement new and rigorous college and career readiness standards and the innovative assessments designed to measure student learning against these standards. The Center on Standards and Assessment Implementation (CSAI) is a federally funded national center charged with focusing research- and evidence-based technical assistance to increase states’ capacity to support their districts and schools in this implementation effort.

CSAI’s theory of action begins with the new college and career readiness standards that provide the framework for classroom instruction and student learning. Research and best practice have shown us that the degree to which there is coherence and alignment among the standards, curricular materials, and instructional strategies used is directly correlated to opportunities for student learning.

The standards also provide the foundation for developing meaningful and effective assessment. The alignment between the standards and assessments is key in determining to whom the assessments are administered and how the data are used. Issues of technical adequacy, including validity (content, construct, predictive, consequential), reliability (measurement precision, stability/consistency, scoring), and fairness (with implications for diverse student populations), are critical to consider in developing, identifying, or evaluating diagnostic, interim, benchmark, and summative assessments. This is especially true as student achievement data is increasingly used as a metric for accountability at the teacher, school, and district levels.

As seen in our theory of action model, there is not only alignment among curriculum and instruction and assessment, but also a continuous feedback loop among the three, as each informs the others to provide a valid and accurate measure of student learning.

Although CSAI does not work directly in classrooms, we apply this model through the lens of supporting the needs of our diverse learning population at the center of our work. This is reflected in the research, technical assistance, and support that is needed at the classroom, school, district, and state levels of decision-making. Our aim is to focus on building capacity, at all levels, in the development of balanced, coherent, and efficient systems of teaching and learning.”

REL Southwest note: CSAI provides one relevant resource, “Interim Assessments In Brief,” available at <https://www.csai-online.org/sites/default/files/resources/4666/InterimAssessmentsBrief.pdf>.

Methods

Keywords and Search Strings

The following keywords and search strings were used to search the reference databases and other sources:

- (“interim assessment” OR “benchmark assessment” OR “interim test” OR “benchmark test” OR “interim testing” OR “benchmark testing”)

Databases and Resources

We searched [ERIC](#) for relevant, peer-reviewed research references. ERIC is a free online library of more than 1.8 million citations of education research sponsored by the Institute of Education Sciences (IES). Additionally, we searched the [What Works Clearinghouse](#).

Reference Search and Selection Criteria

When we were searching and reviewing resources, we considered the following criteria:

- *Date of the publication:* References and resources published from 2005 to present, were included in the search and review.
- *Search priorities of reference sources:* Search priority is given to study reports, briefs, and other documents that are published and/or reviewed by IES and other federal or federally funded organizations, academic databases, including ERIC, EBSCO databases, JSTOR database, PsychInfo, PsychArticle, and Google Scholar.
- *Methodology:* The following methodological priorities/considerations were given in the review and selection of the references: (a) study types—randomized control trials, quasi-experiments, correlational studies, descriptive data analyses, literature reviews, mixed methods analyses, and so forth; (b) target population, samples (representativeness of the target population, sample size, volunteered or randomly selected, and so forth), study duration, and so forth; and (c) limitations, generalizability of the findings and conclusions, and so forth.

This memorandum is one in a series of quick-turnaround responses to specific questions posed by stakeholders in the Southwest Region (Arkansas, Louisiana, New Mexico, Oklahoma, and Texas), which is served by the Regional Educational Laboratory (REL) Southwest at AIR. This memorandum was prepared by REL Southwest under a contract with the U.S. Department of Education’s Institute of Education Sciences (IES), Contract ED-IES-91990018C0002, administered by AIR. Its content does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.